

How Turner Broadcasting can avoid the *Seven Deadly Sins* That  
Can Cause a Data Warehouse Project to Fail

Robert Milton Underwood, Jr.

© 2000

## Table of Contents

<u>Section</u>	<u>Page</u>
Introduction	3
Sin number one	3
Sin number two	5
Sin number three	7
Sin number four	8
Sin number five	12
Sin number six	13
Sin number seven	14
Conclusion	15

## How Turner Broadcasting can avoid the *Seven Deadly Sins* That Can Cause a Data Warehouse Project to Fail

### Introduction

Denis Kosar, in the book, Data Warehouse: Advice from the Experts<sup>1</sup>, discuss what he terms “the seven deadly sins” that militate against effective data warehouse planning and implementation. Companies considering the need for a data warehouse project would be prudent to understand the pitfalls that the seven sins represent so that a fully functioning and successful project can be realized. Each of the seven sins represents a stage or part of the data warehouse building process that, if overlooked, underrepresented, or omitted, would result in an inferior finished product.

#### Sin number one

Sin number one entails the belief that people will use the data warehouse just because it is available for use. Much time and tremendous resources may be utilized to build a data warehouse, but an impressive project will not necessarily attract its intended user audience. What is of initial importance is that a clear set of business objectives must be established that address the reasons for creating a data warehouse and the expected benefits that users will realize.

Kosar suggests that a company should have a business sponsor. Some companies have either ongoing or periodic relationships with other companies in the form of partnerships. In these cases, one company may be creating a data warehouse that could benefit both companies. The other company may “sponsor” the project by committing

---

<sup>1</sup> Bischoff, J. and Alexander, T. 1997. Data Warehouse: Practical Advice from the Experts. Upper Saddle River, NJ: Prentice Hall. 57-70.

needed funds and by lending support for the project by clearly indicating a *need* for the project. The scenario would be slightly different for Turner Broadcasting Sales, Inc. (Turner Sales). Turner Broadcasting System, Inc., the parent company, would most likely be the sponsor for the Turner Sales data warehouse project. The project itself (i.e. the actual data warehouse) would be created by technicians from the Information Technology (IT) department. The importance of having a business sponsor is that both the sponsor and the creator of the data warehouse benefit from its utilization. The success of Turner Sales obviously benefits the parent company, Turner Broadcasting Systems, Inc. The primary benefit of a sponsor is that the sponsor realizes a need for the data warehouse, and therefore, is supportive of the overall effort.

Another issue related to sin number one is that management should not rely on the data warehouse to solve problems. The data warehouse should be recognized as a tool to be used for successfully solving problems as they arise. Turner Sales should therefore continue to use sound management principles when confronted with problems. Then, when the data warehouse is completed, management can determine whether or not information contained within the data warehouse can be of use for their problem-solving procedures.

One final issue related to sin number one is that a company should implement a data warehouse project in phases. Turner Sales is too large of a company to try to incorporate a data warehouse project that would include the total scope of the enterprise. In general, the larger the organization, the more difficult it is to create a full enterprise solution. Turner Sales should create a first phase that is manageable and that can be

clearly evaluated for performance. Once the first phase is deemed successful, then Turner Sales can use their proven methodology to expand the scope of the data warehouse.

### Sin number two

Sin number two deals with the omission of a suitable data warehouse architecture. This sin represents the lack of a complete architectural framework. If the data warehouse project of Turner Sales has a weak architecture, data cannot be loaded, accessed, and delivered optimally. There are at least six issues to consider when planning a suitable architecture. The first issue that Turner Sales should understand is the expected number of end users and functional areas. Do they expect that all possible users will have access to the data warehouse, or are just certain company departments more likely than others to be considered end users?

A second issue for Turner Sales to consider is the diversity of data. *What kind of data will be included should be carefully estimated. Some information may not need to be included.*

The third issue for Turner Sales to consider is the *volume* of data. They need to determine if every single bit of data should be included for every sale or transaction. Including too much data may bog down the system and take up unnecessary storage space.

The fourth issue for Turner Sales to consider is details of the refresh cycle. This will be important for end users so they can know how often to expect the system to be updated. They cannot expect the system to be updated daily, but they can expect something more reasonable, like biweekly or monthly updates.

The fifth issue that Turner Sales must consider with regards to sin number two is to determine the storage ability of their data warehouse. They must make sure that they have a storage capacity that will accommodate anticipated future needs.

The sixth issue for Turner Sales to understand regarding sin number two is with regards to accessibility. They need to determine *who* will be allowed to access the system, and *when* will the system be available for access. Not every user will be granted full access to the system, but all users should have full access to the parts of the system that will enable them to best perform their jobs.

There are three tiers in the data warehouse architecture: (a) information acquisition, (b) information storage, and (c) information delivery. There are concerns with each individual tier. Each tier must be designed well so that Turner Sales will not experience problems with sin number two. The information acquisition layer is responsible for obtaining, refining, and aggregating the data from the existing production or legacy systems. The function of this architecture layer is to provide data consolidation and integration. Standardization of the system is important because it will be easier for Turner Sales to add future requirements.

The information storage tier is the holding level for obtaining data. It holds the data in snapshot, or historical, form. Successful development of the acquisition tier will allow data of higher integrity into the information storage tier.

The information delivery tier will support a common set of presentation and analytical tools. Different views will be available, but Turner Sales must make sure that there is a common look for all of them. Employees may be confused and their data searches may be slowed if they have numerous singular views that they must get used to.

Also, only the actual subset of data that the user requests should be delivered back to the user through a view. Getting unnecessary data could be confusing and may allow the user to create inaccurate inferences from the information obtained.

### Sin number three

Sin number three is the failure to document all assumptions and conflicts early. Turner Sales is a part of a huge company. Because of its size, Turner Sales should make a special effort to list assumptions that they have about the project. These assumptions can then be used as guidelines as they progress with the data warehouse project. Because of conflicts that may arise, the assumptions may need to be modified as the project development progresses. It is usually during the feasibility phase of the project that questions are identified and documented. The answers to these questions, and their implications, are of very high importance to the development of the data warehouse. One question that is usually asked early in the feasibility stage is *how much data should be loaded initially into the data warehouse?* The answer is usually based on the number of expected users and the amount of data they need to have to perform their jobs successfully. Having too much data can be costly and may not be fully utilized.

A second question that is usually asked during the feasibility phase is *what is the level of granularity that is needed?* Responding to this question will determine whether detailed data should be included, summarized data, or both.

A third question that is useful to ask early in the feasibility phase is *how often should the data be refreshed?* The refresh cycle that Turner Sales decides upon may depend on the update cycles of the operational systems that provide the source of the data. If, for example, the update cycle of the operational system were on a monthly basis,

then it would not be necessary to refresh the data in the data warehouse on a weekly basis.

#### Sin number four

Sin number four deals with the abuse of methodology and tools. There may be numerous methodologies and/or sets of tools to use when constructing a data warehouse. Turner Sales will want to maintain a consistent methodology, especially if what they currently have works well.

The data warehouse tools that Turner Sales chooses will depend on their corporate culture, technical environment, and methodology. According to Kosar, there are four groups of tools: (a) analysis tools, (b) development tools, (c) implementation tools, and (d) delivery tools.

During the analysis phase of the project, Turner Sales would use analysis tools when the current operational environment is being studied. Analysis tools assist in the identification of data requirements. Examples of analysis tools that Turner Sales should consider are CASE tools, scanners, and data repositories.

Turner Sales' data warehouse development team would use development tools to assist during code generation for the information acquisition, data cleansing, data integration, and loading of the data warehouse data. Examples of development tools that Turner Sales could use are code generators and data repositories.

Implementation tools assist in the actual cleansing, consolidation and loading of the data warehouse. Examples of implementation tools that Turner Sales could use are data acquisition tools and information storage tools.



Delivery tools would assist Turner Sales' developers in data conversion, data derivation, and data loading. Delivery tools would also assist developers in the reporting for the delivery platform, if there would be a different platform for delivery. Examples of delivery tools include data loader, data glossary, and query and reporting tools.

The three major functions that describe the use of data warehouse tools are gathering, storing, and delivering. During the gathering (a.k.a. acquisition) function, language scanners are used to capture and document database definitions, data files, data elements, and current legacy files.

During the store function, Turner Sales would capture what was scanned from the legacy systems into a data repository or dictionary tool. During the store function, other important data such as data entities, attributes, relationships, and business rules are also contained.

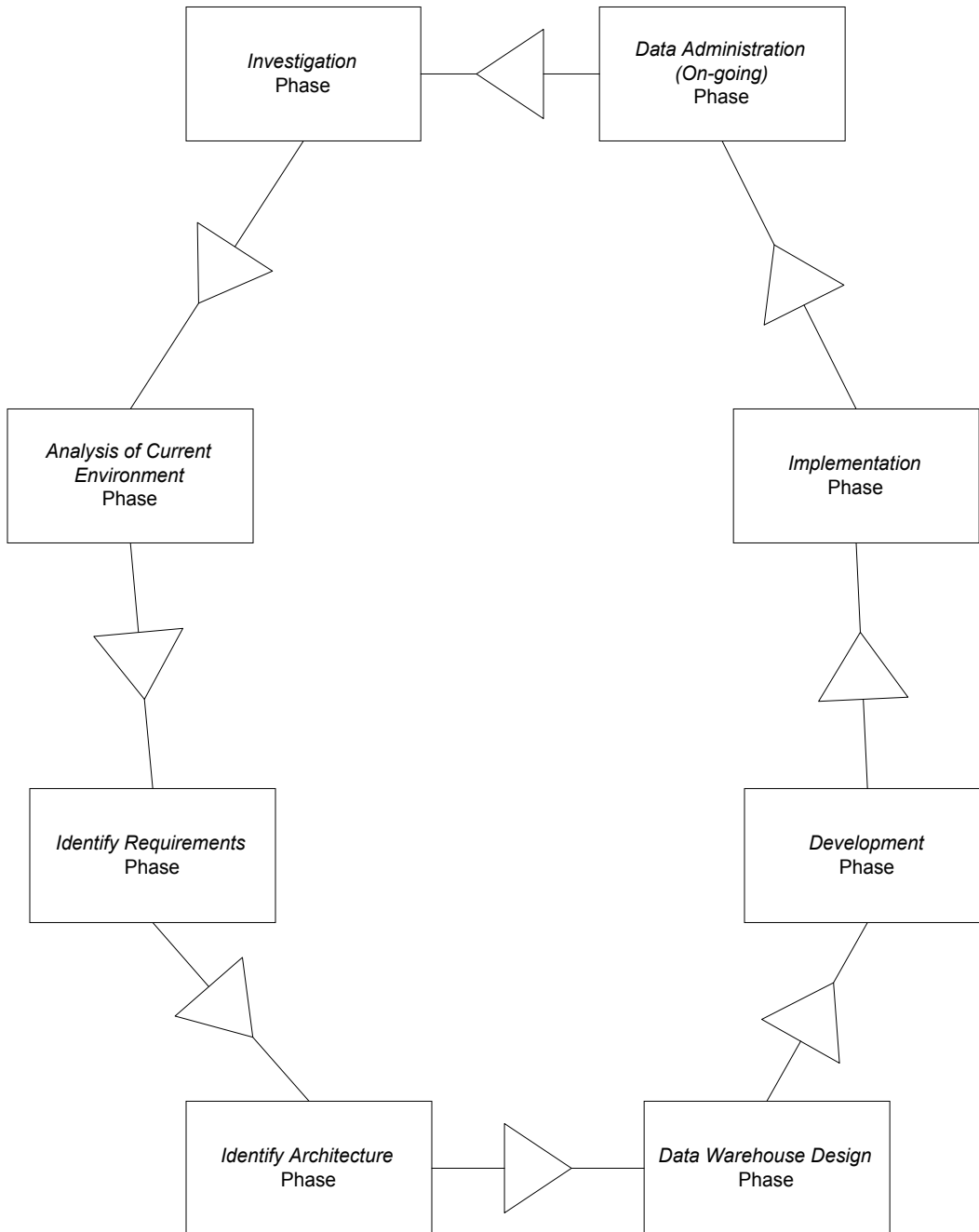
The deliver function will provide the delivery of all of the metadata (i.e. "data about data") to the project developers and eventually to the end users. The deliver function also allows for a user data glossary to assist the business user. The user glossary would contain all definitions that employees at Turner Sales would use within the functional organization.

#### Sin number five

Sin number five deals with the abuse of the data warehouse life cycle. According to Kosar, this sin is actually a failure to realize the difference between the data warehouse life cycle (DWLC) and the system development life cycle (SDLC). There are distinctions between the two life cycle models. Unlike the SDLC, the life cycle for the DWLC is ongoing (see Figure 1). If Turner Sales fails to realize the ongoing nature of the DWLC,

they will be setting themselves up for future problems. The SDLC is used in the development of online transaction processing (OLAP) systems, which are extremely important for organizations. Decisions that the data warehouse offers area used to *support*

Figure 1  
Phases of the Data Warehouse Life Cycle



OLAP development. Turner Sales should be aware of the function of each of the phases of the DWLC, as described below.

Although placed in an ordered list, the stages are actually ongoing and may repeat themselves if another cycle is needed (e.g. for system expansion, as in the case of a corporate merger; or system reengineering, in the case of improvements in technology.) The first phase of the DWLC is the *investigation phase*. During this phase, Turner Sales would identify the need for the data warehouse. A business sponsor is selected during this phase to add support and/or resources. A mission statement for the data warehouse project is developed during this phase.

The second phase of the DWLC is the *analysis of current environment* phase. During this phase, the legacy data is analyzed and documented as necessary. A master file is then made for each application system. A data inventory catalog, list of business definitions, and synonym report are created during this phase.

The third phase of the DWLC is the *identify requirements* phase. During this phase, Turner Sales will determine how the business will use the data warehouse. A requirements document is created during this phase. Note that requirements are needed only for the initial business area for which the data warehouse is being created, and should not be fully developed for the entire enterprise.

The fourth phase of the DWLC is the *identify architecture* phase. The data warehouse architecture is developed by the team from Turner Sales by using the requirements from the previous phase. The team will decide upon a development methodology and a tool kit. Topography for development and implementation is determined during this phase.

The fifth phase of the DWLC is the *data warehouse design* phase. This is when the actual design is produced. Both the logical and physical designs are completed. Each entity is mapped to the corresponding table name during this phase.

The sixth phase of the DWLC is the *development* phase. Turner Sales' development team will develop and test programs for data cleansing, data integration, and data loading. During this phase, the team will finish the data warehouse definitions and the program code. They will also test data and will confer with a sample of users to ascertain how the project is accepted. From these meetings with users will come the framework for quality assurance plans.

The seventh phase of the DWLC is the *implementation* phase. It is during this phase that the fully functional data warehouse is produced. The initial data loads are performed. Other tasks to be completed within this phase are the creation of backup and recovery procedures, definition of user procedures, publishing of training manuals, and actual user queries.

The eighth phase of the DWLC is *ongoing data administration*. During this phase, Turner Sales will monitor data from the feeder systems. Also, their IT department will maintain responsibility for the administration of the metadata. Legacy data from operational systems may change constantly, so update cycles will be scheduled to acquire the changes. The data administration group at Turner Sales will ultimately be responsible for monitoring all changes to the system and data.

If Turner Sales recognizes the importance of each of the eight phases of the DWLC, they will ensure that problems are not encountered with sin number five. The

important thing to remember is that the system is ongoing and will evolve over time as the company evolves.

### Sin number six

Sin number six is failure to properly analyze data and resolve data conflicts. This sin can basically be described as the ignorance concerning the resolution of data conflicts. Different departments within a company may keep records slightly differently. For instance, “*CustomerID*,” “*Customer\_ID*,” “*CustomerNumber*,” and “*Cust\_Number*,” may all refer to the same data field, but different departments may catalog the data differently. Since each department may use their own referencing system, it may be difficult to get all to agree on a change to standardization. Turner Sales must be aware that there is a natural tendency for groups to be protective of their data and the systems that they are used to. All must realize how standardization can benefit the company as a whole. Turner Sales will probably need to review all customer, transaction, and product files to see where inconsistencies reside. If different departments also have different sizes of data fields, then it may be necessary to use the largest data field. For instance, if the Sales Department and the Accounting Department used *Last\_Name* fields that were 13 and 15 spaces long, respectively, it may be practical to choose the 15-space data field as the new standard.

The major tasks that Turner Sales must complete to avoid problems with sin number six are listed below:

1. Identification of key files and systems.
2. Cataloging each field with definitions in a data repository or data dictionary.
3. Outlining, then building a data model.

4. Identification of synonyms concerning data fields. Synonyms, as described during the discussion of sin number five above, are data fields that have the same meanings, but different names.
5. Mapping of all input data fields to business names in the data model.
6. Updating and normalization of the data model as needed for the data warehouse.

#### Sin number seven

Sin number seven is the failure to learn from mistakes, and can also represent disregard of, or inattention to, the principles discussed in any or all of the prior six sins.

Valuable lessons can be learned from mistakes that have occurred in the past or during creation and implementation of the data warehouse. As mistakes are corrected, it is important to document the associated problem(s) and the subsequent solution(s).

Turner Sales should make sure that they take the time to understand how problems occurred, and realize that those problems do not have to be repeated if careful analysis takes place of past, present, and potential problem areas. If they fail to learn from mistakes that have already occurred, then they may fall victim to sin number seven.

Some mistakes may seem very small indeed, and may be noticed and corrected by only one IT specialist. However, it is important that the individual responsible for correcting the problem document the incident so that future employees can avoid the problem altogether. That way, if the IT specialist leaves the firm, there would be some way for future employees to reference the problem(s) that the data warehouse has experienced and their associated solution(s).

## Conclusion

Many companies have fallen victim to one or more of the seven deadly sins during construction of a data warehouse. Since the project can consume many resources and countless hours, it is important that Turner Sales is clear on their objectives at each stage of the development of their data warehouse. In some ways, they may not have as many problems as other companies may have, despite their size. For instance, their plan to have the entire company's application base run on Microsoft® products will allow them a good platform on which to build consistent data. They have recognized that Microsoft® products integrate well. From Microsoft® Office applications to SQL Server, Turner Sales has chosen computer products that will allow them to minimize problems associated with differences in data types. If Turner Sales pays attentions to each of the important points discussed with regards to the seven deadly sins, they should be positioned to have a very successful data warehouse project.